AFRL-SN-WP-TP-2003-111

THE EVALUATION OF SYNTHETIC APERTURE RADAR IMAGE SEGMENTATION ALGORITHMS IN THE CONTEXT OF AUTOMATIC TARGET RECOGNITION

Kefu Xue, Ph.D.

SEPTEMBER 2002

SENSORS DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7318

NOTICE

USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE U.S. GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.

THIS REPORT HAS BEEN REVIEWED BY THE OFFICE OF PUBLIC AFFAIRS (ASC/PA) AND IS RELEASABLE TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS). AT NTIS, IT WILL BE AVAILABLE TO THE GENERAL PUBLIC, INCLUDING FOREIGN NATIONS.

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

/s/

JASON B. GREGGA, Capt
Project Engineer
ATR & Fusion Algorithms Branch
Sensor ATR Technology Division

/s/

DALE E. NELSON, Ph.D.
Chief, ATR & Fusion Algorithms Branch
Sensor ATR Technology Division
Sensors Directorate

/s/

DAVID W. CHANDLER
Lieutenant Colonel, USAF
Deputy, Sensor ATR Technology Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange and does not constitute approval or disapproval of its ideas or findings.

Do not return copies of this report unless contractual obligations or notice on a specific document requires its return.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2002	Conference Paper Preprint	08/15/2001 – 09/10/2002

4. TITLE AND SUBTITLE

THE EVALUATION OF SYNTHETIC APERTURE RADAR IMAGE SEGMENTATION ALGORITHMS IN THE CONTEXT OF AUTOMATIC TARGET RECOGNITION

5a. CONTRACT NUMBER
F33615-01-C-1853

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62204F

6. AUTHOR(S)

Kefu Xue, Ph.D.

5d. PROJECT NUMBER
6095

5e. TASK NUMBER
04

5f. WORK UNIT NUMBER
08

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Electrical Engineering
Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Sensors Directorate
Air Force Research Laboratory
Air Force Materiel Command
Wright-Patterson Air Force Base, OH 45433-7318

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/SNAT

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-SN-WP-TP-2003-111

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Paper was presented at the Optical Science and Technology (SPIE) Conference, Seattle, WA, 7 July 2002.

14. ABSTRACT
Image segmentation is a process to extract and organize information energy in the image pixel space according to a prescribed feature set. It is often a key preprocess in automatic target recognition (ATR) algorithms. In many cases, the performance of image segmentation algorithms will have significant impact on the performance of ATR algorithms. Due to the variations in feature set definitions and the innovations in the segmentation processes, there is large number of image segmentation algorithms existing in the ATR world. The problem is which image segmentation algorithm performs best for an ATR application. There are a number of measures to evaluate the performance of segmentation algorithms, such as Percentage Pixels Same (pps), Partial Directed Hausdorff (pdh), and Complex Inner Product (cip). In the research, we found that the combination of the three measures shows effectiveness in the evaluation of segmentation algorithms against truth data (human master segmentation). However, we don't know what are the impact of those measures in the performance of ATR algorithms that are commonly measured by Probability of detection (P_{Det}), Probability of false alarm (P_{FA}), Probability of identification (P_{ID}), etc. In all practical situations, ATR boxes are implemented without human observer in the loop. The performance of synthetic aperture radar (SAR) image segmentation should be evaluated in the context of ATR rather than human observers.

15. SUBJECT TERMS

Synthetic Aperture Radar (SAR), Automatic Target Recognition (ATR), SAR image segmentation, ATR baseline algorithms, ATR algorithm evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	22	CAPT. Jason B. Gregga
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER *(Include Area Code)* (937) 255-1115 x4386

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

i

Executive Summary

Image segmentation is a process to extract and organize information energy in the image pixel space according to a prescribed feature set. It is often a key preprocess in automatic target recognition (ATR) algorithms. In many cases, the performance of image segmentation algorithms will have significant impact on the performance of ATR algorithms. Due to the variations in feature set definitions and the innovations in the segmentation processes, there is large number of image segmentation algorithms existing in the ATR world. The problem is which image segmentation algorithm performs best for an ATR application. There are a number of measures to evaluate the performance of segmentation algorithms, such as Percentage Pixels Same (pps), Partial Directed Hausdorff (pdh) and Complex Inner Product (cip). In the research, we found that the combination of the three measures shows effectiveness in the evaluation of segmentation algorithms against truth data (human master segmentation). However, we don't know what are the impact of those measures in the performance of ATR algorithms that are commonly measured by Probability of detection (P_{Det}), Probability of false alarm (P_{FA}), Probability of identification (P_{ID}), etc. In all practical situations, ATR boxes are implemented without human observer in the loop. The performance of synthetic aperture radar (SAR) image segmentation should be evaluated in the context of ATR rather than human observers.

In this research, a limited literature search has been conducted. As we expected, there have been no apparent research efforts associated to the evaluation of image segmentation algorithms for ATR in the literature. Most of the image segmentation evaluation methods are limited to the comparison with human observers or truth data. This research project establishes a preliminary segmentation algorithm evaluation suite involving segmentation algorithm performance measures as well as the ATR algorithm performance measures. The suite includes a set of baseline ATR algorithms and standard ATR performance evaluation measures. It provides a quantitative evaluation method to judge which SAR image segmentation algorithm is the best for a particular ATR application. Preliminary experiment results based on some baseline ATR algorithms and a typical image segmentation algorithm using the evaluation suite are tabulated in this report. It is our conclusion that using traditional evaluation measures does not necessarily reflect the performance and impact of image segmentation algorithm towards the ATR performance. Higher in traditional image segmentation performance score does not guarantee the higher in ATR performance score.

A significant portion of the research effort spend on identifying and implementing a suite of baseline ATR algorithms. To our surprise that there is no existing suite of baseline ATR algorithms in the ATR community, even though every one is referring to it. In order to establish a baseline ATR algorithms for testing, we studied an existing software on template matching ATR algorithm developed at AFRL[5] and added two

1

more algorithms to form a baseline ATR algorithm suite. The additional algorithms are conditional Gaussian model based ATR[6] and a SAR ATR system developed in the Lincoln Laboratory[7].

The another outcome from this research effort is a road map leading to further research in this area. Even though, invoking ATR algorithm suite is effective in evaluating the image segmentation algorithms in the context of ATR. There are several problems: a) The image feature space of an image segmentation algorithms does not necessarily match the region of interest of the ATR algorithms in the suite; b) The uncertainty (multiple thresholds and choices of parameters) in the ATR algorithm implementation; c) It is costly every time to invoking multiple ATR algorithms in the evaluation of an image segmentation algorithm. A road map for the further research to overcome those problems is presented in this report.

1 Introduction

Image segmentation is often a key step in an automatic target recognition system. Therefore, the performance of image segmentation algorithm is closely related to the performance of an ATR system. In the ATR world, the image segmentation is also referred to as a pixel-based ATR algorithm that labels the pixels in an image as target pixels (also known as "Point of Interest" (POI)) or non-target pixels. The object-based ATR algorithm[1] creates label to an object-sized group of POIs that is referred to as "Region of Interest" (ROI). It is clear that the performance of object-based ATR should have great dependency on the performance of image segmentation algorithm. Evaluation of the performance of SAR image segmentation also encounters the difficulty that the absolute truth segmentation is often not available. Even the manual segmentations (pseudo-truth) of the same image by different human experts are often dissimilar. That is due to the SAR imagery generated by a radar instrument does not match the ordinary human visual perception. SAR image segmentation requires many trained expert skills and depending on some subjective judgements. In the ATR applications, the SAR image segmentation is done automatically as a part of the ATR system without human in the loop. We think the evaluation should be done in the same automatic fashion. In addition, the segmentation performance should be measured in terms of its application purpose and functionality in an ATR system.

This research effort is trying to establish a research road map leading to a standard image segmentation algorithm evaluation suite in the context of ATR. We have conducted literature search to see if there are some prior work in this area. We have implemented three baseline ATR algorithms which use segmented ROI for object label. Some preliminary experiments were conducted to show that using image segmentation quality metric alone does not necessarily consistent with the quality of ATR performance.

2   Image Segmentation Quality Metric

In the literature search, we did not come cross any prior works on image segmentation metric in the context of ATR problems. There are a number of image segmentation quality metric to provide an objective measure of the performance of image segmentation algorithms by comparing the segmentation results with truth data or master segmented results. In addition, those metric were created based on the segmentation of visual band optical images and not for the feature spaces of SAR images. Some of the image segmentation quality measures have been introduced in the past works [2] [3] [4]. For the convenience of discussion, we make a concise summary of those measures in the following.

Percent Pixels Same (PPS) is to compare two images where all the POIs are classified as "1" and background pixels are classified as "0". PPS measures the percentage of pixels that are classified the same way between two binary images, S_1 and S_2 where S_1 is considered to be truth.

$$pps = \frac{\sum_{i,j} S_1(i,j) S_2(i,j)}{\max\{\sum_{i,j} S_1(i,j), \sum_{i,j} S_2(i,j)\}} \tag{1}$$

When $pps = 1$, it indicates image $S_1 = S_2$ a perfect match. As $pps = 0$, it means a complete miss match. PPS measure over punishes the misalignment of an otherwise perfect segmentation.

Partial Directed Hausdorff (PDH) measures the correctness of the closed boundaries of segmentations. Let $A = \{a_0, a_1, \cdots, a_{N-1}\}$ forms a closed contour of a segmentation in image S_1 and $B = \{b_0, b_1, \cdots, b_{N-1}\}$ is a set of contour vertices of a segmentation in image S_2. The partial directed Hausdorff distance is defined as

$$h_K(A, B) = K^{th}_{a_j \in A}\{\min_{b_i \in B}\{||a_j - b_i||\}\} \tag{2}$$

where K^{th} is the Kth ranked distance. When $K = N - 1$, the partial Hausdorff distance takes the maximum distance in the set A that is referred to as direct Hausdorff distance.

$$pdh(\delta) = \frac{(K|_{h_K(A,B)=\delta}) + 1}{N} \tag{3}$$

where δ is a prescribed threshold and K is the result of solving $h_K(A, B) = \delta$ indicating the total number$(K + 1)$ of vertices within the distance threshold δ. When $pdh(\delta) = 1$, it means the segmentations are matched perfectly. If $pdh(\delta) = 0$, it is a complete mismatch. Again PDH measure over punishes misalignment. In addition, it often requires re–sample the contours to assure that both contours in the comparison have the exactly same number of contour vertices.

Complex Inner Product (CIP) also measures the correctness of the closed boundaries of segmentations. In this case, the Fourier descriptor[2] is used to represent the contours for achieving some level of invariance in contour scaling, rotation and shifting. Let $A(k)$ and $B(k)$ are the Fourier descriptor of contours in images S_1 and S_2 respectively $k = 0, 1, \cdots, N-1$. To avoid instability, an amplitude modulated phase-only filter is used to represent the Fourier descriptor of the truth contour $A(k)$,

$$\frac{\exp(-j\angle\{A(k)\})}{|A(k)| + \varepsilon} \tag{4}$$

where ε is a very small positive constant to avoid division by zero. Since the phase of Fourier descriptor, $\angle\{A(k)\}$, preserves the shape of the contour, the comparison of two contours can be represented by the correlation function,

$$R_{A,B}(n) = \frac{1}{N} \sum_{k=0}^{N-1} B(k) \cdot \frac{\exp(-j\angle\{A(k)\})}{|A(k)| + \varepsilon} e^{\frac{j2\pi}{N}nk}. \tag{5}$$

If two contours match perfectly, $R_{A,B}(n)$ should show a distinctive peak. The CIP measuring the peakness of $R_{A,B}(n)$ is defined as

$$cip = 1 - \frac{1}{\log_{10}\left(\frac{(N-1)\cdot\max_n\{R_{A,B}(n)\}}{\sum_{n=0}^{N-1}|R_{A,B}(n)|-\max_n\{R_{A,B}(n)\}}\right) + 1} \tag{6}$$

where cip ranges from zero to one as one for the perfect match[3].

All the above mentioned measures can be used together to achieve a composite measure. However, there are no clear ways to determine what kind of composite measure should be used under what situation[3]. In intra–algorithmic comparison, the measures are used to compare the segmentation result with the truth data that is often a master segmentation created by a human analyst. If truth data is not available, the segmentation result can be compared with the result generated by another segmentation algorithm using so–called inter–algorithmic comparison[2]. The comparison using these measures are limited to the domain of point of interest and have no link to the performance of ATR algorithms.

3   ATR Algorithm Performance Measures

Object–based ATR algorithms produce label to a ROI that consists segmented group of POIs. ROIs have different categories. For example, ROI could be a group of POIs that are extracted edge components or may be a group of bright pixels that are referred to as target scattering centers. In most of cases, ROI is just a square block of SAR

image pixels that contain target information. Based on the characteristics of ROI, an ATR algorithm creates label or labels and attached them to the ROI. The performance of the ATR algorithms is measured by a confusion matrix[1]. The number of rows

		Decision columns j						
Typical Comfusion Matrix		Detection Accept						Detection Reject
		Target					Non-Target	
		ID Accept				ID Reject		
		Type 1	Type 2	Type 3	Other			
Target	Type 1	M11	M12	M13	M1O	M1R-ID	M1NT	M1R-Det
	Type 2	M21	M22	M23	M2O	M2R-ID	M2NT	M2R-Det
	Type 3	M31	M32	M33	M3O	M3R-ID	M3NT	M3R-Det
	Other	MO1	MO2	MO3	MOO	MOR-ID	MONT	MOR-Det
Non-Target	Type 4	M41	M42	M43	M4O	M4R-ID	M4NT	M4R-Det
	Type 5	M51	M52	M53	M5O	M5R-ID	M5NT	M5R-Det

Figure 1: A typical confusion matrix

depends on the number of test objects. In the ATR algorithm evaluation, a bin of test images of various types of objects is prepared. Each time a test image of certain type of object passing through the ATR algorithm under evaluation, a matrix cell in that row corresponding to the type of object will get an increment. As to which cell (column) getting increment, it depends upon the label decision made by the ATR algorithm, The matrix cell M_{ij} indicates the number of times, a type i object is labeled as type j. For each experimental bin, a confusion matrix will be generated. In order to compare the performance of different ATR algorithms using the same experimental bin or the same ATR algorithm using different experimental bins under various operating conditions, a number summary performance measures will be used in the evaluation. Here is a short list for a number of commonly used summary performance measures[1].

Target Declaration Probability ($P_{tgt-dec}$): This number indicates to what percentage of the test object images, the ATR algorithm can make a decision. The higher

the percentage, the better the performance.

$$P_{tgt-dec} = \frac{\sum\limits_{i=1,2,3,O,4,5} \left(\sum\limits_{j=1,2,3,O,R-ID,NT} M_{ij} \right)}{\sum\limits_{i=1,2,3,O,4,5} \left(\sum\limits_{j=1,2,3,O,R-ID,NT,R-Det} M_{ij} \right)} \tag{7}$$

Detection Probability (P_{Det}): P_{Det} is conditioned only on targets that separates the detection performance evaluation on target objects and non–target objects. Higher P_{Det} means the algorithm has higher ability to positively detect the targets.

$$P_{Det} = \frac{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O,R-ID} M_{ij} \right)}{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O,R-ID,NT} M_{ij} \right)} \tag{8}$$

A similar summary performance measure to P_{Det} is P_s, **Probability of Success** that includes the ability to positively detect non-target objects as well.

$$P_s = \frac{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O,R-ID} M_{ij} \right) + M_{NT,NT}}{\sum\limits_{i=1,2,3,O,4,5} \left(\sum\limits_{j=1,2,3,O,R-ID,NT} M_{ij} \right)} \tag{9}$$

Probability of False Alarm (P_{FA}): It measures the percentage of non–target objects falsely labeled as target objects.

$$P_{FA} = \frac{\sum\limits_{i=4,5} \left(\sum\limits_{j=1,2,3,O,R-ID} M_{ij} \right)}{\sum\limits_{i=4,5} \left(\sum\limits_{j=1,2,3,O,R-ID,NT} M_{ij} \right)} \tag{10}$$

A good algorithm should have lower P_{FA}.

Identification Declaration Probability (P_{ID-dec}): This measure summarize the ability for an ATR algorithm to make decisions on object types.

$$P_{ID-dec} = \frac{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O} M_{ij} \right)}{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O,R-ID} M_{ij} \right)} \tag{11}$$

Identification Probability (P_{ID}): P_{ID} is conditioned on $P_{tgt-dec}$, P_{Det}, P_{ID-dec} and on targets only. It measures the ability for ATR algorithms to positively identify the object types.

$$P_{ID} = \frac{\sum\limits_{i=1,2,3,O} M_{ii}}{\sum\limits_{i=1,2,3,O} \left(\sum\limits_{j=1,2,3,O} M_{ij} \right)} \tag{12}$$

Correct Label Probability (P_{CL}): P_{CL} is conditioned on $P_{tgt-dec}$, P_{Det}, P_{ID-dec}, targets as well as non-targets. It indicates all labeling results even non-target objects labeled as target objects.

$$P_{CL} = \frac{\sum\limits_{i=1,2,3,O} M_{ii}}{\sum\limits_{i=1,2,3,O,4,5} \left(\sum\limits_{j=1,2,3,O} M_{ij} \right)} \tag{13}$$

These are a few summary performance measures that are redefined and clarified by the Air Force Research Laboratory (AFRL), the Comprehensive Performance Assessment of Sensor Exploitation (COMPASE) Center[1]. With the effort by COMPASE, the unified and clearly defined summary performance measures can effectively assist researchers in ATR community to communicate their results and ATR algorithm performance evaluations. We intend to use these measures to evaluate image segmentation algorithm performance in the context of ATR.

4   Image Segmentation Algorithm Performance Measure in the Context of ATR

In the ATR applications, the image segmentation algorithms intend to label the pixels as "point of interest" or "point of no interest" according to a prescribed feature space. Some of the features such as edges, target scattering centers, shadows, etc., are commonly used in ATR algorithms. The purpose of image segmentation in ATR is to label the POIs and group them into a ROI for ATR algorithms to label the type. The ultimate goal of a segmentation algorithm is to provide POIs precisely according to the prescribed feature spaces such that a positive identification can be achieved by an ATR system. Some of the ATR algorithms may not be very sensitive towards the preciseness of POIs and others may well be. With the variations of experimental bins and different mixtures of operating conditions, the performance of different image segmentation algorithms will affect the performance of various ATR algorithms differently. To ensure a good performance image segmentation algorithm is used in an ATR system, the performance has to be evaluated according to or at least correlating to the ultimate performance

measure of the ATR system. The performance of image segmentation algorithms really should be judged in the context of ATR applications.

4.1 Baseline Algorithms

To support this notion, we first have to establish a set of baseline ATR algorithms that uses a similar feature spaces. In this pursuit, we find out that even the core of the ATR community can not agree on what consists of baseline algorithms due to the complexity of the ATR problems. Some of the ATR algorithms are very complicated with so many ad hoc thresholds and parameters to select that it is impossible to reproduce the comparable results as they published. Some other ATR algorithms do not contain enough details in the publication for us to implement. However, we managed to implement three ATR algorithms for conducting our experiments. In those three algorithms, the segmentation algorithm required is not more than just locating a block ROI that contains the POIs of a potential target.

The first algorithm is the classic template matching method. It is a very simple and reliable ATR algorithm that uses the minimum mean square error (MSE) between the received ROI and stored templates to classify the target chip in class and pose[5]. Under the assumption that over a small azimuth angle, the radar signature of the target remains relatively constant, the templates are formed by averaging all the training images in a small interval of azimuth angle at each azimuth angle location. In our experiment, we have completely reproduced the experimental results by Worrell and Parker[5]. We then did a slight modification on the algorithm by removing the mean from all the image chips. It noticeably improves the performance of the algorithms.

$$(c, \theta) = \min_{i,j} \left\{ \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |T(x, y) \cdot (M_{i,j}(x, y) - s(x - x_{opt}, y - y_{opt}))| \right\}_{\substack{i=0,\cdots,P-1 \\ j=1,\cdots,Q-1}} \quad (14)$$

where (c, θ) are class and pose that minimize the error. M_{ij} is the template for the i_{th} pose and j_{th} class. There are total of P number of poses in each class and total of Q number of classes. T is a the binary mask for the brightness of threshold that was the quarter power function in our implementation. $s(x - x_t, y - y_t)$ is the translated SAR image chip (ROI) magnitude.

The second algorithm uses the conditionally Gaussian stochastic signal modeling approach to SAR ATR[6]. In this approach, the radar return is modeled as a complex Gaussian process

$$r = s(\theta, c) + w \quad (15)$$

where s is a complex Gaussian random vector conditioned on (θ, c) with conditional mean $\mu(\theta, c) = 0$ and a diagonal covariance matrix $K(\theta, c)$. w is a zero mean Gaussian

8

random noise with covariance matrix $N_0\mathbf{I}$. The conditional covariance matrix of the radar return r can be found as

$$E\{(rr^T|\theta, c\} = K(\theta, c) + N_0\mathsf{I}. \tag{16}$$

The Log likelihood function becomes

$$l(r|\theta, c) = \sum_i \left[-\ln(K_{i,j}(\theta, c) + N_0) - \frac{|r_i|^2}{K_{i,j}(\theta, c) + N_0} \right]. \tag{17}$$

The only unknown in evaluating $l(r|\theta, c)$ is the values of $K_{i,j}(\theta, c) + N_0$ which will be estimated from the training data. Under the assumption that the variance of each pixel is nearly constant within small intervals, $W_k = \left[\frac{2\pi K}{N_w} - \frac{d}{2}, \frac{2\pi K}{N_w} + \frac{d}{2} \right]$, of the posing angle $\theta_k = \frac{2\pi K}{N_w}$, the variance of i_{th} pixel is estimated from SAR testing images of class c as

$$\hat{\sigma}_i^2(\theta_k, c_l) = \frac{1}{N_k} \sum_{\theta \in W_k} |r_i(\theta, c_l)|^2, \qquad 1 \le l \le t, \qquad 1 \le k \le N_w \tag{18}$$

where N_k is the total number of training images within the azimuth interval W_k and t is the number of classes in the database. The classification of the ROI is performed using Bayesian approach by selecting the target class $c(r)$ such that the $P(c|r)$ is maximized.

$$c_{Bayes}(r) = \max_c \sum_k \sum_j P(r|\theta_k, s_j, c) \tag{19}$$

where s_j is all possible object locations.

The third baseline algorithm is developed by the Lincoln Laboratory[7]. The algorithm uses a three stage processes, detector, discriminator, and classifier to label the type of object. The detector is a two–parameter CFAR detector

$$\frac{s(x, y) - \mu_c}{\sigma_c} > K_{CFAR} \tag{20}$$

where $s(x, y)$ is the amplitude of pixel under the test, μ_c and σ_c are mean and standard deviation of the clutter inside the boundary stencil, and K_{CFAR} is a threshold chosen to control the false alarm rate. The boundary stencil is simply a area defined around each test pixel. The discriminator process each ROI produced from the detector and rejects any ROI that does not contain man made objects. The classifier stage consists of a Mean Squared Error template matching classifier..

$$\varepsilon = \frac{\sum_{i=1}^N (R_i - T_i)^2}{N} \tag{21}$$

where N is the number pixels in reference template, R_i are pixels in the normalized reference template and T_i are pixels in the processed test image.

The template matching algorithm has been implemented at AFRL by Michael Bryant in Matlab. We just made some miner changes (remove image mean) to generate best ATR results with MSTAR data. Based on the template matching Matlab code, we implemented the other two algorithms to form a baseline ATR algorithm suite. All these algorithms are only require rectangular shaped ROI segments.

4.2 Preliminary Evaluation Suite

With the baseline algorithms in place, the evaluation algorithms can be evaluated in the context of ATR. The block diagram of evaluation suite is shown in figure 2. Since
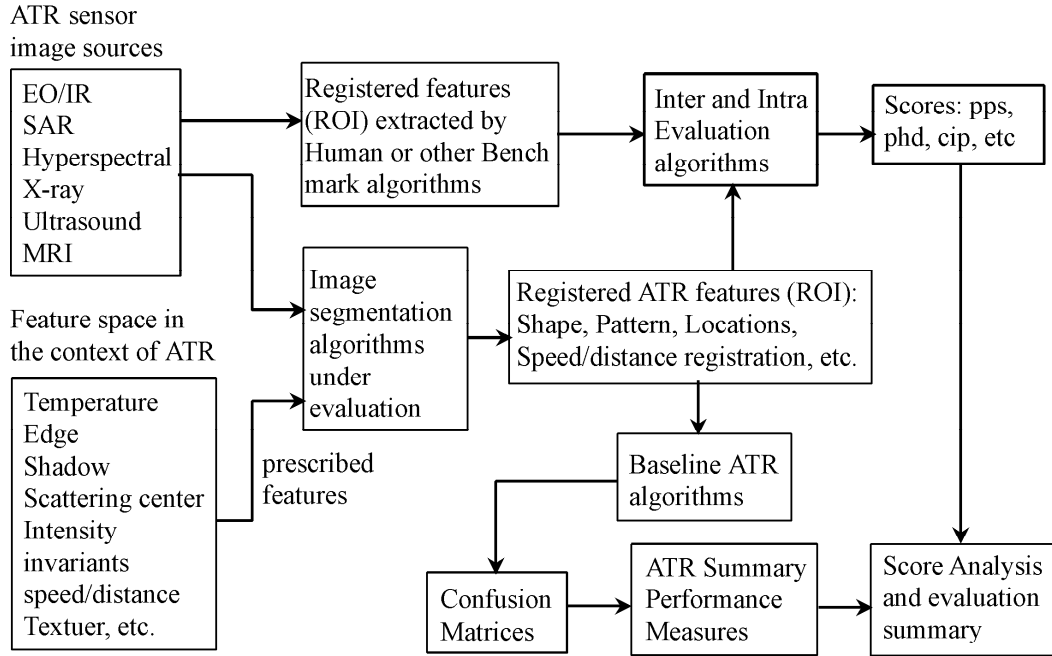
Figure 2: The evaluation suite for image segmentation algorithms in the context of ATR

the scores from image segmentation evaluation (*pps*, *pdh*, and *cip*) and all the scores for ATR summary performance measures are normalized from 0 to 1. The score analysis can be done by various rank orders of the score vector S.

$$S = [pps \quad pdh \quad cip \vdots P_{tgt-dec} \quad P_{Det} \quad P_s \quad (1 - P_{FA}) \quad P_{ID-dec} \quad P_{ID} \quad P_{CL}] \quad (22)$$

Some of the typical rank orders such as maximum value,

$$sc_{\text{max}} = \max_i \{S(i)\}, \tag{23}$$

median

$$sc_{mdn} = \text{median}\{S(i)\} \tag{24}$$

and weighted mean

$$sc_{av} = \sum_i [w(i) \cdot S(i)] \tag{25}$$

where $w(i)$ are selected weighting values with $\sum_i w(i) = 1$. If some of the scores in score vector S are not needed in the evaluation report, then the corresponding weight can be set to zero. A evaluation score use the sum of three top ranked scores in score vector S,

$$sc_{\text{max 3}} = \frac{1}{3} \sum_{\text{max}\{3\}} S(i) \tag{26}$$

We can also use only a subset of the score vector for the algorithm evaluation. For example, the evaluation uses the partial sum of identification scores

$$sc_{ps-id} = (P_{ID-dec} + P_{ID} + P_{CL})/3 \tag{27}$$

We did some preliminary evaluation experiments using the public released Moving and Stationary Target Acquisition and Recognition (MSTAR) SAR image data. The test uses two targets (BMP–2 and T–72) and one confuser (BTR–70) that is considered not a target in this experiment.

The first experiment uses the template matching baseline algorithm. If the ROI is segmented tightly with a size of 64–by–64, the confusion matrix is

	BMP–2	T–72	Non–target
BMP-2	534	21	32
T–72	17	481	84
BTR70	108	9	79

$$(28)$$

and the corresponding partial score vector is

$$S = [P_{Dec} = 0.9008 \quad P_s = 0.9683 \quad (1 - P_{FA}) = 0.4931 \quad P_{ID} = 0.9639 \quad P_{CL} = 0.8675]. \tag{29}$$

The confusion matrix for a segmentation not so tightly with a size of 80–by–80 is

	BMP–2	T–72	Non–target
BMP-2	556	0	31
T–72	0	497	85
BTR70	122	2	72

$$(30)$$

and the partial score vector is

$$S = [P_{Dec} = 0.9008 \quad P_s = 0.9624 \quad (1 - P_{FA}) = 0.3673 \quad P_{ID} = 1 \quad P_{CL} = 0.8946].$$
(31)

The score comparison list

	$sc_{\max}$	sc_{mdn}	sc_{av}	$sc_{\max 3}$	sc_{ps-id}
64–by–64	0.9683	0.9008	0.83872	0.9443	0.9157
80–by–80	1	0.9008	0.8250	0.9544	0.9473
(32)

tells us that the segmentation algorithm segment the target image chip not so tight to the boundary of target pixels (ROI) is better for the template matching ATR performance in majority of the scores.

On the other hand, the experiments on the conditionally Gaussian baseline algorithm show different conclusion. For the segmentation size of 64–by–64, the confusion matrix is

	BMP–2	T–72	Non–target
BMP-2	347	50	58
T–72	67	443	100
BTR70	120	80	50
(33)

and the partial score vector is

$$S = [P_{Dec} = 0.8516 \quad P_s = 0.8986 \quad (1 - P_{FA}) = 0.2 \quad P_{ID} = 0.8710 \quad P_{CL} = 0.7136].$$
(34)

When the segmentation size of 48–by–48 is used, the confusion matrix becomes

	BMP–2	T–72	Non–target
BMP-2	508	29	50
T–72	24	493	66
BTR70	107	84	5
(35)

and the partial score vector is

$$S = [P_{Dec} = 0.9009 \quad P_s = 0.9051 \quad (1 - P_{FA}) = 0.0255 \quad P_{ID} = 0.9497 \quad P_{CL} = 0.8040].$$
(36)

The score comparison matrix

	$sc_{\max}$	sc_{mdn}	sc_{av}	$sc_{\max 3}$	sc_{ps-id}
64–by–64	0.8986	0.8516	0.7070	0.8737	0.7927
48–by–48	0.9497	0.9009	0.7170	0.9186	0.8791
(37)

clearly shows the tight segmentation helps ATR algorithm performance.

As to the scores like *pps*, *pdh*, and *cip*, they would indicate higher score, in general, for the segmentations that are tighter to the ROI. However, it is not necessary the reflection of ATR performance. The experiments support the notion that the evaluation suite to examine the performance of SAR image segmentation should be conducted in the context of ATR performance.

5 Current Problems in Evaluation and Road Map to Future Study

In this report, we have presented a evaluation suite to examine the performance of SAR image segmentation algorithms in the context of ATR performance. The preliminary experiments show that the evaluation in the context of ATR does show some advantages. The composite scores from the evaluation suite do reflect the ATR performance. However, it should be noted that there are still many problems associated with this preliminary image segmentation evaluation suite.

5.1 Problems in current evaluation test

The first problem is that the cost associated with the evaluation suite is much higher than just the test of image segmentation algorithm without invoking the ATR algorithm. Furthermore, the implementation difficulty on various ATR algorithms is another high cost item. Secondly due to the high complexity of ATR algorithms, there may be many uncertainties on the implementation and on the selection of parameters (such as threshold values) of an ATR algorithm. The uncertainties will affect the scores of ATR performance. It is quite often that one implementation performs differently than the other implementation. In order to find out the score change due to the difference in image segmentation algorithms not the difference in the implementation versions of ATR algorithm, we need to calibrate the ATR algorithms in the evaluation suite. For many complicated ATR algorithms, it is very hard to achieve. In our experiments, we just simply to adjust the parameters in the ATR algorithms to achieve their best performance for a given image segmentation. It is also costly and time consuming to achieve that. The third problem is again about efficiency. The task of image segmentation algorithms is to discover all the point of interest based on prescribed features and group them together forming a region of interest. To access the performance of image segmentation algorithms, we only need to have several ROI processed. On the other hand, the objective of ATR algorithms is to label each ROI according to the features (models) of target objects. In the experiment of testing ATR algorithm performance, we often have multiple ROI associated to one object under different operating conditions. With multiple objects in a test, we are using much more information than we needed to judge the performance

of an image segmentation algorithm duo to that the ATR performance evaluation process is invoked.

5.2   Road map to future study

Current research only concludes that the performance of image segmentation algorithms must be judged in the context of ATR. However, to make the evaluation process practical, further research is needed to improve the efficiency of the performance evaluation process.

The image segmentation algorithms under the evaluation always have the same feature space as the ATR algorithm that receives the segmented ROI from them. What we need to find out is a metric or a set of specifications to measure the quality of ROI to an ATR algorithm. The higher quality of ROI, the higher performance of the ATR algorithm. For example, the ATR algorithm using conditional Gaussian model requires a ROI with very tight segmentation boundary due to the fact that ATR performance is based on how close the probability density function (PDF) of an observed object to the PDF of a trained model. Tighter segmentation around an object makes ROI containing more pixels belong to one object. Therefore, the PDF of an observed object becomes closer to the PDF of a trained model. On the other hand, for the template matching method, the ROI of a template model contains both target object pixels and background pixels. An overly tighter segmentation can reduce the accuracy in the matching process. The point is different ATR algorithms have different requirements and different quality metric for the ROI's segmented by image segmentation algorithms. For an efficient image segmentation evaluation system, we should first generate a quality metric of ROI for a targeted ATR algorithm. Based on the quality metric, image segmentation algorithms can be evaluated and compared in the context of a specific ATR algorithm. In many cases, a quality metric can be shared by several ATR baseline algorithms that have the similar requirements on the ROI. Figure 3 illustrates the idea of an efficient image segmentation evaluation suite in the context of ATR.

The first challenge that we are facing is to generate quality metric of ROI that should be closely correlate to the performance score of ATR algorithms. This can be viewed as to establish an input requirements and specifications for any ATR algorithms. We understand this has not been a standard process for now. However, it could be a standard process for the future development of any object–based ATR algorithms. For those classical object–based ATR algorithms, we need to develope and test their quality metric of ROI. The quality metric developed in the study can be served as standard models for any new projects in ATR algorithm development. The initial development and test of quality metric of ROI will be costly and time consuming due to the need of invoking a complete ATR algorithm evaluation. However, with the quality metric of ROI, any future evaluation of image segmentation algorithms will be very efficient. It simply evaluates the ROI generated by an image segmentation algorithm against the

ATR sensor
image sources

| EO/IR
SAR
Hyperspectral
X-ray
Ultrasound
MRI |

Image data →

prescribed
features

Image
segmentation
algorithms
under
evaluation

Feature space in
the context of ATR

| Temperature
Edge
Shadow
Scattering center
Intensity
invariants
speed/distance
Textuer, etc. |

Registered
ATR features
(ROI):
Shape, Pattern,
Locations,
Speed,
distance etc.

Quality
metric
calculation
and Score
vector
generator

Score Analysis
and Evaluation
summary

Quality
Metric of
ROI (I)

Baseline ATR
algorithm 1

Baseline ATR
algorithm 2

Baseline ATR
algorithm 3

Quality
Metric of
ROI (II)

Baseline ATR
algorithm 4

Baseline ATR
algorithm 5
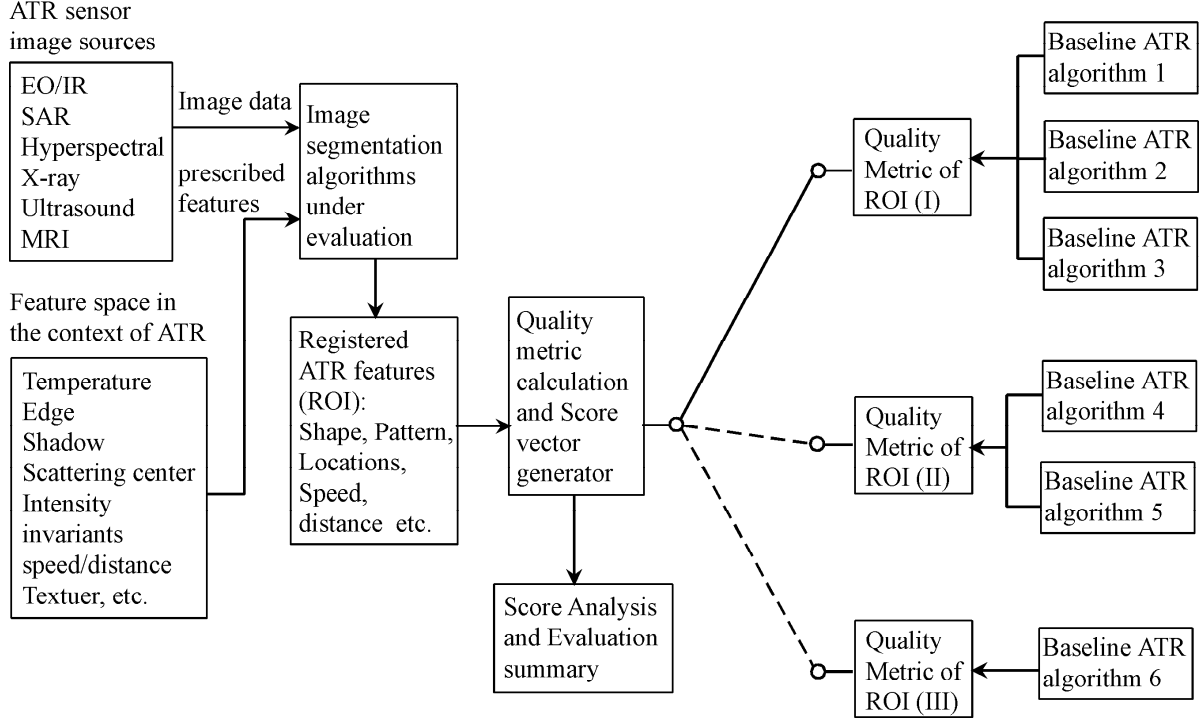
Quality
Metric of
ROI (III)

Baseline ATR
algorithm 6

Figure 3: An efficient image segmentation evaluation suite in the context of ATR

quality metric of ROI and generates scores and evaluation summary. Therefore, the second challenge is to develop a score system using the quality metric of ROI. But for this part, we have many prior work to reference.

As a summary of conclusion, we complete the research by finding that for ATR application, the image segmentation algorithms have to be evaluated in the context of ATR. While, as shown in our experiments, ATR performance measures can be used as an indicator to the performance of image segmentation algorithm, the evaluation process is very inefficient and high in cost. To over come the problem of inefficiency, we planned a future research road map for improvement. The future research in establishing a quality metric of ROI for every object–based ATR algorithms can effectively improve the efficiency of image segmentation algorithm evaluation in the context of ATR. The concept is very simple that the object–based ATR algorithms have to have a specific measurement metric to quantify the quality of its inputs, ROI. The higher the quality measure, the better the ATR performance. And, at the same time, the output (ROI) from the image segmentation algorithm used in the ATR should be measured using the same quality metric of ROI established by the ATR algorithm.

References

[1] Timothy D. Ross, Lori A. Westerkamp, Ronald L. Dilsavor, and John C. Mossing, "Performance Measures for Summarizing Confusion Matrices — The AFRL COMPASE Approach," Proceedings of SPIE Vol. 4727, April 2002, Paper No. 4727-31.

[2] Gregory J. Power, "Multimetric Inter-Algorithmic Evaluation of SAR Image Segmentation Algorithms," Proceedings of SPIE, Vol. 4382, Algorithms for Synthetic Aperture Radar Imagery VIII, April 2001, pp. 330–337.

[3] Jason B. Gregga, Steven C. Gustafson, and Gregory J. Power, "An Objective Evaluation of Four SAR Image Segmentation Algorithms," Proceedings of SPIE, Vol. 4382, Algorithms for Synthetic Aperture Radar Imagery VIII, April 2001, pp. 346–357.

[4] Gregory J. Power and Abdul A.S. Awwal, "Optoelectronic Complex Inner Product for Evaluating Quality of Image Segmentation," Proceedings of SPIE 4114, July 2000.

[5] Stephen W. Worrell and S. Parker, "Class Separability Assessments and MSE Algorithm robustness," Proceedings of SPIE, Vol. 4382, Algorithms for Synthetic Aperture Radar Imagery IV, 1997, pp. 294–304.

[6] J.A. O'Sullvan, M.D. Devore, V. Kedia, and M.I. Miller, "SAR ATR Performance Using Conditionally Gaussian Model," IEEE Transactions on Aerospace and Electronic Systems, Vol. 37(1), pp. 91–105, Jan. 2001.

[7] Leslie M. Novak, Shawn D. Halversen, Gregory J. Owirka, and Margarita Hiett, "Effects of Polarization and Resolution on the Performance of a SAR Automatic Target Recognition System," The Lincoln Laboratory Journal, 6(1): 49–67, 1995.